POLYNOMIAL SPLINE CONFIDENCE BANDS FOR REGRESSION CURVES

Jing Wang and Lijian Yang

University of Illinois at Chicago, Michigan State University

Abstract: Asymptotically exact and conservative confidence bands are obtained for nonparametric regression function, using piecewise constant and piecewise linear spline estimation, respectively. Compared to the pointwise confidence interval of Huang (2003), the confidence bands are inflated by a factor of $\{\log(n)\}^{1/2}$, with the same width order as the Nadaraya-Watson bands of Härdle (1989), the local polynomial bands of Xia (1998) and Claeskens and Van Keilegom (2003). Simulation experiments provide strong evidence that corroborates with the asymptotic theory. The linear spline bands has been applied to identify an appropriate polynomial trend for fossil data.

Short Running Title. SPINE CONFIDENCE BANDS

Key words and phrases: Brownian bridge, B spline, knots, nonparametric regression, quantile transformation

1 . Introduction

For two decades, nonparametric regression has been widely applied to biostatistics, econometrics, engineering and geography, due to its flexibility in modelling complex relationships among variables by "letting the data speak for themselves". Two popular nonparametric smoothing techniques are local polynomial/kernel and polynomial spline.

The kernel type estimators, namely the Nadaraya-Watson and the local polynomial estimator, are based on the locally weighted averaging. The polynomial spline estimators are "global" in terms of implementation, as a single least square procedure leads to the ultimate function estimate over the entire data range, see Stone (1994). In terms of pointwise asymptotics, of course, both kernel and spline type estimators are local in nature, see Fan and Gijbels (1996) and Huang (2003).

The fidelity of a nonparametric regressor is measured in terms of its rate of convergence to the unknown regression function. The convergence rate can be pointwise, least square or uniform. For kernel type estimators, rates of convergence of all three types have been established by Claeskens and Van Keilegom (2003), Fan and Gijbels (1996), and Mack and Silverman (1982), to name a few. For polynomial splines, least squares rates of convergence have been obtained by Stone (1994), while pointwise convergence rates and asymptotic distribution have been recently established in Huang (2003). Confidence band for polynomial spline regression, however, is available only under the restriction of homoscedastic normal errors, see Zhou, Shen and Wolfe (1998).

In this paper, we present confidence bands of univariate regression function based on polynomial spline smoothing. We assume that observations $\{(X_i, Y_i)\}_{i=1}^n$ and unobserved errors $\{\varepsilon_i\}_{i=1}^n$ are i.i.d. copies of (X, Y, ε) satisfying the regression model

$$Y = m(X) + \sigma(X)\varepsilon, \tag{1.1}$$

where the joint distribution of (X, ε) satisfies Assumption (A4) in Section 2. The unknown mean and standard deviation functions $m(x)$ and $\sigma(x)$, defined on interval $[a, b]$, need not to be of any specific form. If the data actually follows a polynomial regression model, $m(x)$ would be a polynomial and $\sigma(x)$, a constant.

Confidence band has been obtained for kernel type estimators of $m(x)$, see Claeskens and Van Keilegom (2003), Hall and Titterington (1988), Hädle (1989), and Xia (1998). These are computationally intensive as the kernel estimator requires solving an optimization problem at every point. In contrast, it is enough to solve only one such problem to get the polynomial spline estimator. The greatest advantages of polynomial spline estimation are its simplicity of implementation and fast computation. Hence, it is desirable from a theoretical as well as a practical point of view to have a confidence band for polynomial spline estimators.

We organize our paper as follows. In Section 2 we state our main results on confidence bands constructed from (piecewise) constant/linear splines. In Section 3 we provide further insights into the error structure of spline estimators. Section 4 describes the actual steps to implement the confidence bands. Section 5 reports findings in an extensive simulation study and the testing of polynomial trend hypothesis for the fossil data using linear spline band. Section 6 concludes. All technical proofs are contained in Appendices A and B.

2 . Main Results

To introduce the spline functions, divide the finite interval $[a,b]$ into $(N+1)$ subintervals $J_j = [t_j, t_{j+1})$, $j = 0, ..., N-1$, $J_N = [t_N, b]$. A sequence of equally-spaced points $\{t_j\}_{j=1}^{N}$, called interior knots, are given as

$$t_0 = a < t_1 < \cdots < t_N < b = t_{N+1}, t_j = a + jh, \ j = 0, 1, ..., N+1,$$

in which $h = (b-a)/(N+1)$ is the distance between neighboring knots. We denote by $G^{(p-2)} = G^{(p-2)}[a,b]$ the space of functions that are polynomials of degree $(p-1)$ on each J_j and has continuous $(p-2)$th derivative. For example, $G^{(-1)}$ denotes the space of functions that are constant on each J_j, and $G^{(0)}$ denotes the space of functions that are linear on each J_j and continuous on $[a,b]$.

In what follows, $\|\cdot\|_\infty$ denotes the supremum norm of a function r on $[a,b]$, i.e. $\|r\|_\infty = \sup_{x\in[a,b]} |r(x)|$, and the moduli of continuity of a continuous function r on $[a,b]$ is denoted as $\omega(r,h) = \max_{x,x'\in[a,b],|x-x'|\le h} |r(x) - r(x')|$. One has $\lim\limits_{h\to 0} \omega(r,h) = 0$ by the uniform continuity of r on a compact interval $[a,b]$.

An asymptotic exact (conservative) $100(1-\alpha)\%$ confidence band for the unknown $m(x)$ over interval $[a,b]$ consists of an estimator $\hat{m}(x)$ of $m(x)$, lower and upper confidence limit $\hat{m}(x) - l_n(x)$, $\hat{m}(x) + l_n(x)$ at every $x \in [a,b]$ such that

$$\begin{aligned} \lim\nolimits_{n\to\infty} P\{m(x) \in \hat{m}(x) \pm l_n(x), \forall x \in [a,b]\} &= 1-\alpha, \text{ exact}, \\ \liminf\nolimits_{n\to\infty} P\{m(x) \in \hat{m}(x) \pm l_n(x), \forall x \in [a,b]\} &\ge 1-\alpha, \text{ conservative}. \end{aligned}$$

Our approach is to get the following polynomial spline estimator based on data $\{(X_i, Y_i)\}_{i=1}^n$ drawn from model (1.1)

$$\hat{m}_p(x) = \underset{g\in G^{(p-2)}[a,b]}{\operatorname{argmin}} \sum\nolimits_{i=1}^n \{Y_i - g(X_i)\}^2, p = 1, 2, \tag{2.1}$$

and then construct the error bound function $l_n(x)$ around this spline estimator. The technical assumptions we need are as follows:

(A1) *The regression function* $m(\cdot) \in C^{(p)}[a,b]$, $p = 1, 2$.

(A2) *The density function* $f(\cdot)$ *of* X *is continuous and positive on interval* $[a,b]$. *The standard deviation function* $\sigma(\cdot) \in C[a,b]$ *has bounded variation and positive lower bound on* $[a,b]$.

(A3) *The subinterval length* $h \sim n^{-1/(2p+1)}$. *I.e., the number of interior knots* $N \sim n^{1/(2p+1)}$.

(A4) *The joint distribution* $F(x,\varepsilon)$ *of random variables* (X,ε) *satisfies:*

(a) *The error is a white noise:* $E(\varepsilon\,|X = x) = 0,\ E(\varepsilon^2\,|X = x) = 1.$

(b) *There exists a positive value* $\delta > 1/p$ *and finite positive* M_δ *such that* $E|\varepsilon|^{2+\delta} < M_\delta$ *and* $\sup_{x\in[a,b]} E\left(|\varepsilon|^{2+\delta}\,|X = x\right) < M_\delta$.

Assumptions (A1)-(A3) are the same as in Huang (2003), while (A4) is the same as (C2) (a) of Mack and Silverman (1982). All are typical for nonparametric regression, with (A1), (A2) and (A4) weaker than the counterparts in Härdle (1989).

To properly define the confidence bands, we introduce some additional notations. For any $x \in [a, b]$, define its location and relative position indices $j(x), \delta(x)$ as

$$j(x) = j_n(x) = \min\{[(x-a)/h], N\}, \delta(x) = \left\{x - t_{j(x)}\right\}/h. \tag{2.2}$$

Since any x is between two consecutive knots, it is clear that $t_{j_n(x)} \leq x < t_{j_n(x)+1},\ 0 \leq \delta(x) < 1, \forall x \in [a, b)$, and $\delta(b) = 1$. Denote by $\|\phi\|_2$ the theoretical L^2 norm of a function ϕ on $[a, b]$, $\|\phi\|_2^2 = E\left\{\phi^2(X)\right\} = \int_a^b \phi^2(x) f(x)\,dx$, and the empirical L^2 norm as $\|\phi\|_{2,n}^2 = n^{-1}\sum_{i=1}^n \phi^2(X_i)$. Corresponding inner products are defined by

$$\langle\phi, \varphi\rangle = \int_a^b \phi(x)\varphi(x) f(x)\,dx = E\{\phi(X)\varphi(X)\}, \langle\phi, \varphi\rangle_n = n^{-1}\sum\nolimits_{i=1}^n \phi(X_i)\varphi(X_i).$$

for any L^2-integrable functions ϕ, φ on $[a, b]$. Clearly $E\langle\phi, \varphi\rangle_n = \langle\phi, \varphi\rangle$.

Although the truncated power basis is used in implementation (see Section 4), it is more convenient to work with the B-spline basis for theoretical analysis. The B-spline basis of $G^{(-1)}$, the space of piecewise constant splines, are indicator functions of intervals $J_j, b_{j,1}(x) = I_j(x) = I_{J_j}(x), 0 \leq j \leq N$. The B-spline basis of $G^{(0)}$, the space of piecewise linear splines, are $\{b_{j,2}(x)\}_{j=-1}^N$

$$b_{j,2}(x) = K\left(\frac{x - t_{j+1}}{h}\right), j = -1, 0, ..., N, \text{ for } K(u) = (1 - |u|)_+ .$$

Define the rescaled B-spline basis $\{B_{j,p}(x)\}_{j=1-p}^N$ for $G^{(1-p)}$

$$B_{j,p}(x) \equiv b_{j,p}(x)\,\|b_{j,p}\|_2^{-1}, 1 - p \leq j \leq N, p = 1, 2.$$

Obviously all the rescaled basis function will have the theoretical norm 1.

To express the estimator $\hat{m}_p(x)$ based on the standardized basis $\{B_{j,p}(x)\}_{j=1-p}^N$, we introduce the following vectors in R^n for $p = 1, 2$

$$\mathbf{Y} = (Y_1, ..., Y_n)^T, \mathbf{B}_{j,p}(\mathbf{X}) = \{B_{j,p}(X_1), ..., B_{j,p}(X_n)\}^T, j = 1 - p, ..., N.$$

The definition of $\hat{m}_p(x)$ in (2.1) entails that $\hat{m}_p(x) \equiv \sum_{j=1-p}^{N} \hat{\lambda}_{j,p} B_{j,p}(x)$ where the coefficients $\left\{\hat{\lambda}_{1-p,p}, ..., \hat{\lambda}_{N,p}\right\}^T$ are solutions of the following least squares problem

$$\left\{\hat{\lambda}_{1-p,p}, ..., \hat{\lambda}_{N,p}\right\}^T = \underset{R^{N+p}}{\operatorname{argmin}} \sum\nolimits_{i=1}^{n} \left\{Y_i - \sum\nolimits_{j=1-p}^{N} \lambda_{j,p} B_{j,p}(X_i)\right\}^2. \tag{2.3}$$

Typically those are the solutions of the normal equations

$$\left(\langle B_{j,p}, B_{j',p}\rangle_n\right)_{j,j'=1-p}^{N} \left(\hat{\lambda}_{j,p}\right)_{j=1-p}^{N} = \left(n^{-1} \sum\nolimits_{i=1}^{n} B_{j,p}(X_i) Y_i\right)_{j=1-p}^{N}.$$

It is straightforward to see that $\langle B_{j,p}, B_{j',p}\rangle \equiv 0, |j-j'| \geq p$, thus the inner product matrix on the left side of the normal equation is diagonal for the constant B spline basis ($p = 1$), and tridiagonal for the linear B spline basis ($p = 2$). According to Lemma 3.1, it is approximated by its deterministic version whose inverse has explicit formula given in Section 4.

For $p = 2$, this inverse matrix S and its 2×2 diagonal submatrices $\{S_j, 0 \leq j \leq N\}$ are expressed as

$$S = (s_{j,j'})_{j,j'=-1}^{N} = (\langle B_{j,2}, B_{j',2}\rangle)^{-1}, S_j = \begin{pmatrix} s_{j-1,j-1} & s_{j-1,j} \\ s_{j,j-1} & s_{j,j} \end{pmatrix}. \tag{2.4}$$

The width of the confidence bands depends on the heteroscedastic variance function. Define

$$\sigma_{n,1}^2(x) = \frac{\int_{I_{j(x)}} \sigma^2(v) f(v)\, dv}{n \|b_{j(x),1}\|_2^2}, \sigma_{n,2}^2(x) = \sum_{j,j',l,l'=-1}^{N} \frac{B_{j',2}(x) B_{l',2}(x) s_{jj'} s_{ll'} \sigma_{jl}}{n} \tag{2.5}$$

with $j(x)$ defined in (2.2), and $s_{ll'}$ in (2.4), and

$$(\sigma_{jl})_{j,j'=-1}^{N} = \Sigma = \left\{\int \sigma^2(v) B_{j,2}(v) B_{l,2}(v) f(v)\, dv\right\}_{j,j'=-1}^{N}. \tag{2.6}$$

These $\sigma_{n,p}^2(x)$ are shown in Lemmas A.4, B.4 to be the pointwise variance functions of $\hat{m}_p(x), p = 1, 2$.

We now state our main results in the next two theorems.

Theorem 1 *Under Assumptions (A1)-(A4), if $p = 1$, then an asymptotic $100(1-\alpha)\%$ exact confidence band for $m(x)$ over interval $[a, b]$ is*

$$\hat{m}_1(x) \pm \sigma_{n,1}(x) \{2\log(N+1)\}^{1/2} d_n, \tag{2.7}$$

in which $\sigma_{n,1}(x)$ is given in (2.5) and can be replaced by $\sigma(x)\{f(x)nh\}^{-1/2}$, according to (A.7) in Lemma A.4, and

$$d_n = 1 - \{2\log(N+1)\}^{-1}\left[\log\left\{-\frac{\log(1-\alpha)}{2}\right\} + \frac{\log\log(N+1) + \log 4\pi}{2}\right]. \tag{2.8}$$

The confidence band in Theorem 1 is superior to the connected error bar of Hall and Titterington (1988) in two aspects: we treat random instead of equally-spaced designs, and by applying the strong approximation theorem of Tusnády (1977), our confidence band is asymptotically exact rather than conservative. The upcrossing results (Theorem 4) used in the proof of Theorem 1 is also different from that used in Bickel and Rosenblatt (1973), Hädle (1989), and Rosenblatt (1976).

Theorem 2 *Under Assumptions (A1)-(A4), if* $p = 2$*, then an asymptotic* $100\,(1-\alpha)\,\%$ *conservative confidence band for* $m\,(x)$ *over interval* $[a, b]$ *is*

$$\hat{m}_2\,(x) \pm \sigma_{n,2}\,(x)\,\{2\log\,(N+1) - 2\log\alpha\}^{1/2}\,, \tag{2.9}$$

where $\sigma_{n,2}\,(x)$ *is as in (2.5), replaceable by* $\sigma\,(x)\,\{2f\,(x)\,nh/3\}^{-1/2}\,\mathbf{\Delta}^T\,(x)\,S_{j(x)}\mathbf{\Delta}\,(x)$ *according to Lemma B.4, and by* $\sigma\,(x)\,\{2f\,(x)\,nh/3\}^{-1/2}\,\mathbf{\Delta}^T\,(x)\,\Xi_{j(x)}\mathbf{\Delta}\,(x)$ *according to Lemma B.3.*

Theorem 2 on linear confidence band bears no similarity to the local polynomial bands in Claeskens and Van Keilegom (2003) and Xia (1998) except the width of the band being of the same order $n^{-1/5}\,(\log n)^{1/2}$. The asymptotic variance function $\sigma^2_{n,2}\,(x)$ of $\hat{m}_2\,(x)$ in (2.5) is a special unconditional version of equation (6.2), in Huang (2003). Thus, the linear band localized at any given point x, is only a factor of $(\log n)^{1/2}$ wider than the pointwise confidence interval of (2003).

3 . Error Decomposition

In this section, we break the estimation error $\hat{m}_p\,(x) - m\,(x)$ into a bias term and a noise term. To understand this decomposition, we begin by discussing the spline space $G^{(p-2)}$ and the representation of the spline estimator $\hat{m}_p\,(x)$ in (2.1).

We note first the uniform convergence of the empirical inner product to the theoretical counterparts.

Lemma 3.1 *Under Assumptions (A2) and (A3), as* $n \to \infty$

$$A_{n,1} = \sup_{0\le j\le N}\left|\|B_{j,1}\|^2_{2,n} - 1\right| = O_p\left(\sqrt{n^{-1}h^{-1}\log\,(n)}\right), \tag{3.1}$$

$$A_{n,2} = \sup_{g_1,g_2\in G^{(0)}}\left|\frac{\langle g_1, g_2\rangle_n - \langle g_1, g_2\rangle}{\|g_1\|_2\,\|g_2\|_2}\right| = O_p\left(\sqrt{n^{-1}h^{-1}\log\,(n)}\right). \tag{3.2}$$

We write $\mathbf{Y}$ as the sum of a signal vector $\mathbf{m}$ and a noise vector $\mathbf{E}$

$$\mathbf{Y} = \mathbf{m} + \mathbf{E}, \mathbf{m} = \{m(X_1), ..., m(X_n)\}^T, \mathbf{E} = \{\sigma(X_1)\varepsilon_1, ..., \sigma(X_n)\varepsilon_n\}^T.$$

Projecting this relationship into the linear space spanned by $G_n^{(p-2)} = \{\mathbf{B}_{j,p}(\mathbf{X})\}_{j=1-p}^N$, a subspace of R^n, one gets

$$\hat{\mathbf{m}}_p = \{\hat{m}_p(X_1), ..., \hat{m}_p(X_n)\}^T = \mathrm{Proj}_{G_n^{(p-2)}} \mathbf{Y} = \mathrm{Proj}_{G_n^{(p-2)}} \mathbf{m} + \mathrm{Proj}_{G_n^{(p-2)}} \mathbf{E}.$$

Correspondingly in the space $G^{(p-2)}$ of spline functions, one has

$$\hat{m}_p(x) = \tilde{m}_p(x) + \tilde{\varepsilon}_p(x) \tag{3.3}$$

$$\tilde{m}_p(x) = \sum\nolimits_{j=1-p}^N \tilde{\lambda}_{j,p} B_{j,p}(x), \tilde{\varepsilon}_p(x) = \sum\nolimits_{j=1-p}^N \tilde{a}_{j,p} B_{j,p}(x). \tag{3.4}$$

The vectors $\left\{\tilde{\lambda}_{1-p,p}, ..., \tilde{\lambda}_{N,p}\right\}^T$ and $\{\tilde{a}_{1-p,p}, ..., \tilde{a}_{N,p}\}^T$ are solutions to (2.3) with Y_i replaced by $m(X_i)$ and $\sigma(X_i)\varepsilon_i$ respectively.

We cite next two important results from de Boor (2001) and Huang (2003).

Theorem 1 *There exists an absolute constant $C_p > 0, p \geq 1$ such that for every $m \in C^{(p)}[a,b]$, there exists a function $g \in G^{(p-2)}[a,b]$ such that*

$$\|g - m\|_\infty \leq C_p \left\|\omega\left(m^{(p-1)}, h\right)\right\|_\infty h^{p-1} \leq C_p \left\|m^{(p)}\right\|_\infty h^p.$$

Theorem 2 *There exists an absolute constant $C_p > 0, p \geq 1$ such that for any $m \in C^{(p)}[a,b]$ and the function $\tilde{m}_p(x)$ defined in (3.4),*

$$\|\tilde{m}_p(x) - m(x)\|_\infty \leq C_p \inf_{g \in G^{(p-2)}} \|g - m\|_\infty = O_p(h^p). \tag{3.5}$$

According to Theorem 2, the bias term $\tilde{m}_p(x) - m(x)$ is of order $O_p(h^p)$. Hence the main hurdle of proving Theorems 1 and 2 is the noise term $\tilde{\varepsilon}_p(x)$. This is handled by the next two propositions.

Proposition 3.1 *With $\sigma_{n,1}(x)$ given in (2.5), the process $\sigma_{n,1}(x)^{-1}\tilde{\varepsilon}_1(x), x \in [a,b]$ is almost surely uniformly approximated by a Gaussian process $U(x), x \in [a,b]$ with covariance structure*

$$EU(x)U(y) = \sum_{j=0}^N I_j(x) \cdot I_j(y) = \delta_{j(x),j(y)}, \forall x, y \in [a,b],$$

where δ_{jl} is the Kronecker symbol, i.e., $\delta_{j,l} = 1$ if $j = l$ and 0 otherwise.

Proposition 3.2 *For a given $0 < \alpha < 1$,and $\sigma_{n,2}(x)$ as given in (2.5)*

$$\liminf_{n\to\infty} P\left[\sup_{x\in[a,b]} \left|\sigma_{n,2}^{-1}(x)\tilde{\varepsilon}_2(x)\right| \leq \{2\log(N+1) - 2\log\alpha\}^{1/2}\right] \geq 1 - \alpha. \tag{3.6}$$

We state next the strong approximation theorem of Tusnády (1977), which will be used later in the proof of Lemmas A.6 and B.6, key steps in proving Proposition 3.1 and Proposition 3.2.

Theorem 3 *Let* $U_1, \ldots, U_n$ *be i.i.d. r.v.'s on the* 2 *-dimensional unit square with* $P(U_i < \mathbf{t}) = \lambda(\mathbf{t}), \mathbf{0} \le \mathbf{t} \le \mathbf{1}$, *where* $\mathbf{t} = (t_1, t_2)$ *and* $\mathbf{1} = (1, 1)$ *are* 2*-dimensional vectors,* $\lambda(\mathbf{t}) = t_1 t_2$. *The empirical distribution function* $F_n^u(\mathbf{t})$ *based on sample* $(U_1, \ldots, U_n)$ *is defined as* $F_n^u(\mathbf{t}) = n^{-1} \sum_{i=1}^n I_{\{U_i < \mathbf{t}\}}$ *for* $\mathbf{0} \le \mathbf{t} \le \mathbf{1}$. *The* 2*-dimensional Brownian bridge* $B(\mathbf{t})$ *is defined by* $B(\mathbf{t}) = W(\mathbf{t}) - \lambda(\mathbf{t}) W(\mathbf{1})$ *for* $\mathbf{0} \le \mathbf{t} \le \mathbf{1}$, *where* $W(\mathbf{t})$ *is a* 2*-dimensional Wiener process. Then there is a version of* $F_n^u(\mathbf{t})$ *and* $B(\mathbf{t})$ *such that*

$$P\left[\sup_{\mathbf{0} \le \mathbf{t} \le \mathbf{1}} \left| n^{1/2} \left\{ F_n^u(\mathbf{t}) - \lambda(\mathbf{t}) \right\} - B(\mathbf{t}) \right| > (C \log n + x) \frac{\log n}{n^{1/2}} \right] < K e^{-\lambda x} \tag{3.7}$$

holds for all x, *where* C, K, λ *are positive constants.*

The well-known Rosenblatt quantile transformation is denoted as

$$(X', \varepsilon') = M(X, \varepsilon) = \left\{ F_X(x), F_{\varepsilon|X}(\varepsilon|x) \right\}, \tag{3.8}$$

which produces random variables X' and ε' with independent and identical uniform distribution on the interval $[0, 1]$. This transformation had been used in, for instance, Bickel and Rosenblatt (1973), Härdle (1989). Substituting the vector $\mathbf{t} = (t_1, t_2)$ in Theorem 3 with (X', ε'), and the stochastic process $n^{1/2} \{ F_n^u(\mathbf{t}) - \lambda(\mathbf{t}) \}$ with

$$Z_n \left\{ M^{-1}(x', \varepsilon') \right\} = Z_n(x, \varepsilon) = \sqrt{n} \left\{ F_n(x, \varepsilon) - F(x, \varepsilon) \right\}, \tag{3.9}$$

where $F_n(x, \varepsilon)$ denotes the empirical distribution of (X, ε), then (3.7) implies that there exists a version of 2-dimensional Brownian bridge B such that

$$\sup_{x, \varepsilon} \left| Z_n(x, \varepsilon) - B\{M(x, \varepsilon)\} \right| = O\left(n^{-1/2} \log^2 n\right), \text{w.p.1.} \tag{3.10}$$

The next result on upcrossing probability is from Leadbetter, Lindgren, and Rootzén (1983), Theorem 1.5.3. It plays the role of Theorem A1 in Bickel and Rosenblatt (1973) or Theorem C in Rosenblatt (1976).

Theorem 4 *If* $\xi_1, ..., \xi_n$ *are i.i.d. standard normal r.v.'s, then for* $M_n = \max\{\xi_1, ..., \xi_n\}, \tau \in R$, *as* $n \to \infty$

$$P\{a_n(M_n - b_n) \le \tau\} \to \exp\left(-e^{-\tau}\right), P\{|M_n| \le \tau/a_n + b_n\} \to \exp\left(-2e^{-\tau}\right),$$

where $a_n = (2 \log n)^{1/2}, b_n = (2 \log n)^{1/2} - \frac{1}{2} (2 \log n)^{-1/2} (\log \log n + \log 4\pi)$.

4 . Implementation

In this section, we describe the procedures to implement the confidence bands in Theorems 1 and 2. Our codes are written in XploRe for convenience of using kernel smoothing, see Härdle, Hlávka and Klinke (2000).

Given any sample $\{(X_i, Y_i)\}_{i=1}^n$ from model (1.1), we use $\min(X_1, ..., X_n)$ and $\max(X_1, ..., X_n)$ respectively as the endpoints of interval $[a, b]$. Minor adjustments could be made for outliers. The number of interior knots is taken to be $N = \left[c_1 n^{1/(2p+1)}\right] + c_2$, where c_1 and c_2 are positive integers. Since explicit formula of coverage probability does not exist for the bands, there is no optimal method to select (c_1, c_2). In simulation, the simple choice of 5 for c_1 and 1 for c_2 seems to work well, so these are set as default values.

The least squares problem in (2.1) can be solved via the truncated power basis $\{1, x, ..., x^{p-1}, (x - t_j)_+^{p-1}, j = 1, ..., N\}$. In other words

$$\hat{m}_p(x) = \sum_{k=0}^{p-1} \hat{\gamma}_k x^k + \sum_{j=1}^{N} \hat{\gamma}_{j,p} (x - t_j)_+^{p-1}, p = 1, 2 \tag{4.1}$$

where the coefficients $\left\{\hat{\gamma}_0, ..., \hat{\gamma}_{p-1}, \hat{\gamma}_{1,p}, ..., \hat{\gamma}_{N,p}\right\}^T$ are solutions of the following least squares problem

$$\left\{\hat{\gamma}_0, ..., \hat{\gamma}_{N,p}\right\}^T = \underset{R^{N+p}}{\operatorname{argmin}} \sum_{i=1}^{n} \left\{Y_i - \sum_{k=0}^{p-1} \gamma_k X_i^k - \sum_{j=1}^{N} \gamma_{j,p} (X_i - t_j)_+^{p-1}\right\}^2.$$

When constructing the confidence bands, one needs to evaluate the functions, $\sigma_{n,p}^2(x)$ in (2.5) differently for the exact and conservative bands, and the description is separated into two subsections. For both cases, following Lemmas A.4, B.4, one estimates the unknown functions $f(x)$ and $\sigma^2(x)$ and then plugs in these estimates, the same approach taken in Hall and Titterington (1988), Hädle (1989), and Xia (1989). This is analogous to using $\overline{X} \pm 1.96 \times s_n/\sqrt{n}$ instead of $\overline{X} \pm 1.96 \times \sigma/\sqrt{n}$ as a large sample 95% confidence interval for a normal population mean μ, where the sample standard deviation s_n is a plugin substitute for the unknown population standard deviation σ.

Let $\widetilde{K}(u) = 15\left(1 - u^2\right)^2 I\{|u| \leq 1\}/16$ be the quartic kernel, s_n =the sample standard deviation of $(X_i)_{i=1}^n$ and

$$\hat{f}(x) = n^{-1} \sum_{i=1}^{n} h_{\text{rot},f}^{-1} \widetilde{K}\left(\frac{X_i - x}{h_{\text{rot},f}}\right), h_{\text{rot},f} = (4\pi)^{1/10} \left(\frac{140}{3}\right)^{1/5} n^{-1/5} s_n \tag{4.2}$$

where $h_{\text{rot},f}$ is the rule-of-thumb bandwidth of Silverman (1986). Define next matrices $\mathbf{Z}_p = \{Z_{1,p}, .., Z_{n,p}\}^T, p = 1, 2$ with $Z_{i,p} = \{Y_i - \hat{m}_p(X_i)\}^2$ and

$$\mathbf{X} = \mathbf{X}(\mathbf{x}) = (X_1 - x, ..., X_n - x)^T, \mathbf{W} = \mathbf{W}(x) = \operatorname{diag}\left\{\widetilde{K}\left(\frac{X_i - x}{h_{\text{rot},\sigma}}\right)\right\}_{i=1}^{n},$$

where $h_{\mathrm{rot},\sigma}$ =the rule-of-thumb bandwidth of Fan and Gijbels (1996) based on data $(X_i, Z_{i,p})_{i=1}^n$. Then one defines the following estimators of $\sigma^2(x)$

$$\hat{\sigma}_p^2(x) = \left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}\mathbf{Z}_p, p = 1, 2. \tag{4.3}$$

The following uniform consistency results are provided in Bickel and Rosenblatt (1973) and Fan and Gijbels (1996)

$$\max_{p=1,2}\sup_{x\in[a,b]}\left|\hat{\sigma}_p(x) - \sigma(x)\right| + \sup_{x\in[a,b]}\left|\hat{f}(x) - f(x)\right| = o_p(1). \tag{4.4}$$

4.1 Implementing the exact band

The function $\sigma_{n,1}(x)$ is approximated by either one of the following, with $\hat{f}(x)$ and $\hat{\sigma}_1(x)$ defined in (4.2) and (4.3), $j(x)$ defined in (2.2)

$$\begin{aligned}\hat{\sigma}_{n,1}(x,1) &= \hat{\sigma}_1\left(t_{j(x)}\right)\hat{f}^{-1/2}\left(t_{j(x)}\right)n^{-1/2}h^{-1/2}\\ \hat{\sigma}_{n,1}(x,2) &= \hat{\sigma}_1(x)\hat{f}^{-1/2}(x)n^{-1/2}h^{-1/2},\end{aligned}$$

where the additional parameter value 1 or 2 indicating the estimation at each value x or at the nearest left knot. Since $\sup_{x\in[a,b]}\left|x - t_{j(x)}\right| \le h \to 0$, as $n \to \infty$, (4.4) entails that both of the bands below are asymptotically exact with $\hat{m}_1(x)$ given in (4.1) and d_n in (2.8)

$$\hat{m}_1(x) \pm \hat{\sigma}_{n,1}(x, \mathrm{opt})\{2\log(N+1)\}^{1/2}d_n, \mathrm{opt} = 1, 2. \tag{4.5}$$

4.2 Implementing the conservative band

According to Lemma B.3, for $0 \le j \le N$, the matrix Ξ_j approximates matrix S_j uniformly. Hence both of the bands below are asymptotically conservative, with $\hat{m}_2(x)$ given in (4.1)

$$\hat{m}_2(x) \pm \hat{\sigma}_{n,2}(x, \mathrm{opt})\{2\log(N+1) - 2\log\alpha\}^{1/2}, \mathrm{opt} = 1, 2, \tag{4.6}$$

where the function $\sigma_{n,2}(x)$ in (2.5) for the linear band is estimated consistently by either one of the next two formulae

$$\begin{aligned}\hat{\sigma}_{n,2}(x,1) &= \left\{\mathbf{\Delta}^T(x)\,\Xi_{j(x)}\mathbf{\Delta}(x)\right\}^{1/2}\hat{\sigma}_2\left(t_{j(x)}\right)\left\{\frac{2}{3}\hat{f}\left(t_{j(x)}\right)nh\right\}^{-1/2},\\ \hat{\sigma}_{n,2}(x,2) &= \left\{\mathbf{\Delta}^T(x)\,\Xi_{j(x)}\mathbf{\Delta}(x)\right\}^{1/2}\hat{\sigma}_2(x)\left\{\frac{2}{3}\hat{f}(x)\,nh\right\}^{-1/2},\end{aligned}$$

with $j(x)$ defined in (2.2), and $\hat{f}(x)$ and $\hat{\sigma}_2(x)$ defined in (4.2) and (4.3), $\mathbf{\Delta}(x)$ and Ξ_j defined as follows:

$$\mathbf{\Delta}(x) = \begin{pmatrix} c_{j(x)-1}\{1-\delta(x)\} \\ c_{j(x)}\delta(x)\end{pmatrix}, c_j = \begin{cases}\sqrt{2} & j = -1, N \\ 1 & j = 0, ..., N-1\end{cases}, \tag{4.7}$$

$$\Xi_j = \begin{pmatrix} l_{j+1,j+1} & l_{j+1,j+2} \\ l_{j+2,j+1} & l_{j+2,j+2}\end{pmatrix}, j = 0, 1, ..., N, \tag{4.8}$$

with terms $l_{ik}, |i-k| \leq 1$ defined through the following matrix inversion

$$M_{N+2} = \begin{pmatrix} 1 & \sqrt{2}/4 & & & & 0 \\ \sqrt{2}/4 & 1 & 1/4 & & & \\ & 1/4 & 1 & \ddots & & \\ & & \ddots & \ddots & 1/4 & \\ & & & 1/4 & 1 & \sqrt{2}/4 \\ 0 & & & & \sqrt{2}/4 & 1 \end{pmatrix}_{(N+2)\times(N+2)} = (l_{ik})^{-1}_{(N+2)\times(N+2)}, \qquad (4.9)$$

and computed via (4.12), (4.13), and (4.14) given below.

In order to calculate the matrix M^{-1}_{N+2}, which is needed for (4.8), we introduce two theorems from matrix theory, equation (43) in Gantmacher and Klein (1960) and Theorem 4.5 in Zhang (1999).

Theorem 1 *For a symmetric Jacobi matrix J given as follows*

$$J = \begin{pmatrix} a_1 & b_1 & & 0 \\ b_1 & \ddots & \ddots & \\ & \ddots & \ddots & b_{N+1} \\ 0 & & b_{N+1} & a_{N+2} \end{pmatrix}_{(N+2)\times(N+2)},$$

its inverse matrix $J^{-1} = (l_{ik})_{(N+2)\times(N+2)}$ satisfies

$$l_{i,k} = \psi_i \chi_k, i \leq k, l_{i,k} = \psi_k \chi_i, k \leq i, \qquad (4.10)$$

where

$$\psi_i = \frac{(-1)^i \det\left(J_{(1,\ldots,i-1)}\right) b_i b_{i+1} \cdots b_{N+1}}{\det(J)}, \chi_k = \frac{(-1)^k \det\left(J_{(k+1,\ldots,N+2)}\right)}{b_k b_{k+1} \cdots b_{N+1}}, \qquad (4.11)$$

and $J_{(1,\ldots,i-1)}$ is defined as the upper left $(i-1)\times(i-1)$ submatrix of J, $\det(J)$ is the determinant of matrix J, while $J_{(k+1,\ldots,N+2)}$ is the corresponding lower right $(N+2-k)\times(N+2-k)$ submatrix.

Theorem 2 *For a tridiagonal matrix given as*

$$T_N = \begin{pmatrix} a & b & & 0 \\ c & a & \ddots & \\ & \ddots & \ddots & b \\ 0 & & c & a \end{pmatrix}_{N\times N}, N \geq 1,$$

if $a^2 \neq 4bc$, then the determinant of T_N is

$$\det T_N = \frac{\alpha^{N+1} - \beta^{N+1}}{\alpha - \beta}, \alpha = \frac{a + \sqrt{a^2 - 4bc}}{2}, \beta = \frac{a - \sqrt{a^2 - 4bc}}{2}.$$

Now we let

$$z_1 = \frac{2+\sqrt{3}}{4}, z_2 = \frac{2-\sqrt{3}}{4}, \theta = \frac{z_2}{z_1} = \left(2-\sqrt{3}\right)^2 = 7-4\sqrt{3}, \tag{4.12}$$

and apply Theorems 1 and 2 to obtain

$$l_{11} = l_{N+2,N+2} = \frac{8z_1^2\left(1-\theta^{N+1}\right) - z_1\left(1-\theta^N\right)}{8z_1^2\left(1-\theta^{N+1}\right) - 2z_1\left(1-\theta^N\right) + 8\left(1-\theta^{N-1}\right)},$$

$$l_{i,i} = \frac{\left\{8z_1\left(1-\theta^{N+2-i}\right) - \left(1-\theta^{N+1-i}\right)\right\}\left\{8z_1\left(1-\theta^{i-1}\right) - \left(1-\theta^{i-2}\right)\right\}}{(z_1-z_2)\left\{64z_1^2\left(1-\theta^{N+1}\right) - 16z_1\left(1-\theta^N\right) + 64\left(1-\theta^{N-1}\right)\right\}} \tag{4.13}$$

for $2 \leq i \leq N+1$ and

$$l_{12} = l_{N+1,N+2} = \frac{\left(-2\sqrt{2}\right)z_1\left(1-\theta^N\right) - \left(1-\theta^{N-1}\right)}{8z_1^2\left(1-\theta^{N+1}\right) - 2z_1\left(1-\theta^N\right) + 8\left(1-\theta^{N-1}\right)},$$

$$l_{i,i+1} = \frac{\left\{8z_1\left(1-\theta^{N+1-i}\right) - \left(1-\theta^{N-i}\right)\right\}\left\{8z_1\left(1-\theta^{i-1}\right) - \left(1-\theta^{i-2}\right)\right\}}{(-4)(z_1-z_2)\left\{64z_1^2\left(1-\theta^{N+1}\right) - 16z_1\left(1-\theta^N\right) + 64\left(1-\theta^{N-1}\right)\right\}} \tag{4.14}$$

for $2 \leq i \leq N$. By the symmetry of matrix M_{N+2}, the lower diagonal entries are $l_{i+1,i} = l_{i,i+1}, \forall i = 1, ..., N+1$.

5 . Examples

5.1 Simulation example

To illustrate the finite-sample behavior of our confidence bands, we simulate data from model (1.1), with $X \sim U[-1/2, 1/2]$, and

$$m(x) = \sin(2\pi x), \sigma(x) = \sigma_0 \frac{100-\exp(x)}{100+\exp(x)}, \varepsilon \sim N(0,1). \tag{5.1}$$

The noise level $\sigma_0 = 0.2, 0.5$ while sample size $n = 100, 200, 500, 10000$. Confidence level $1-\alpha = 0.99, 0.95$. Tables 1 and 2 contain the coverage probabilities as the percentage of coverage of the true curve at all data points by the confidence bands in (4.5) and (4.6), over 500 replications of sample size n. We have also computed the coverage probabilities of the confidence bands in (2.7) by plugging in the true value of density function $f(x) = I_{[-1/2,1/2]}(x)$ and the variance function $\sigma(x)$ in (5.1). These bands are called "oracle bands" as they use quantities that are unknown but for "oracles"; whereas the bands in (4.5) are called "estimated bands".

In Table 1 the surprise is that all four bands have the same coverage at noise level 0.5. At noise level 0.2, the performance of all four becomes much closer with sample sizes increasing, whereas for small sample sizes the oracle bands are slightly better. In Table 2, the coverage percentages show positive confirmation of Theorem 2. At sample size 200, regardless of noise level, both of the two candidate bands in (4.6) achieve at least 95.6% and 90% for confidence level $1 - \alpha = 0.99, 0.95$ respectively.

From both tables, it is obvious that larger sample size guarantees improved coverage, with reasonable coverage achieved at moderate sample sizes. The linear band outperforms the constant band, corroborating with the theory. The noise level has more influence to the constant bands than the linear ones.

For the linear bands, we have also carried out simulation for sample size $n = 10000$ and opt $= 1$. Regardless of the noise level, the coverage is always 99.4% for $\alpha = 0.01$ and 97.6% for $\alpha = 0.05$, both higher than the nominal coverage of 99% and 95%, consistent with their conservative definitions. Remarkably, it takes merely 88 minutes to run 500 simulations with sample size as large as 10000 on a Pentium 4 PC. This is extremely fast considering that nonparametric regression is done without WARPing, see Härdle, Hlávka and Klinke (2000).

The graphs in Figure 4.6 are created based on two samples of size 100 and 500 respectively, each with four types of symbols: points (data), center thin solid line (true curve), center dashed line (the estimated curve), upper and lower thick solid line (confidence bands). In all figures, the confidence bands of $n = 500$ are thinner and fits better than those of $n = 100$.

Table 1: Constant bands coverage probabilities (500 replications).

noise level	**sample size** n	**confidence level**	**estimated bands**	**oracle bands**
0.2	100	0.99	0.476 (0.458)	0.606 (0.606)
		0.95	0.256 (0.246)	0.438 (0.436)
	200	0.99	0.704 (0.708)	0.802 (0.802)
		0.95	0.454 (0.456)	0.532 (0.532)
	500	0.99	0.826 (0.834)	0.832 (0.832)
		0.95	0.462 (0.456)	0.468 (0.468)
0.5	100	0.99	0.618 (0.618)	0.618 (0.618)
		0.95	0.504 (0.504)	0.504 (0.504)
	200	0.99	0.860 (0.860)	0.860 (0.860)
		0.95	0.716 (0.716)	0.716 (0.716)
	500	0.99	0.932 (0.932)	0.932 (0.932)
		0.95	0.802 (0.802)	0.802 (0.802)

Table 2: Linear spline bands coverage probabilities (500 replications).

noise level	**sample size** n	**confidence level** 0.99	**confidence level** 0.95
	100	0.900 (0.896)	0.816 (0.814)
0.2	200	0.956 (0.962)	0.902 (0.904)
	500	0.990 (0.988)	0.954 (0.958)
	100	0.904 (0.904)	0.822 (0.814)
0.5	200	0.956 (0.960)	0.900 (0.902)
	500	0.990 (0.988)	0.956 (0.960)

5.2 Fossil data example

The fossil data reflects global climate millions of years ago through ratios of strontium isotopes found in fossil shell, it was studied by Chaudhuri and Marron (1999) to detect the structure via kernel smoothing. The corresponding penalized spline fits is provided in Ruppert, Wand and Carroll (2003). In this section we test the polynomial form of the fossil data regression curve. The null hypothesis is $H_0 : m(x) = \sum_{k=1}^{d} a_k x^k$, with polynomial degree $d = 2, 3, 5, 6$. The response Y is the strontium isotopes ratio after linear transformation, $Y = 0.70715 + \text{ratio} * 10^{-5}$, since all the values are very close to 0.707, while the predictor X is the fossil shell age in million years.

In Figure 2, the center dotted line is the linear spline fit. The upper/lower thin lines represent linear bands implemented according to (4.6). The solid line is the least square polynomial fit with degrees d. Clearly, the oversmoothed quadratic null curve ($d = 2$) is rejected at significance level 0.01 since it is not totally covered by the confidence bands with confidence level 0.99. When $d = 3, 5$ the null solid curves can capture the big dip at the range of $110 - 115$ million years old, it is still not a good fit. Thus both null parametric models H_0 are rejected at the level 0.01. While in the case $d = 6$, all significant features are shown in the null polynomial curve, the relative high ratio before 105 million years old, the substantial dip around 115 million years old, and the relative flat stage between 95 and 105. Given a 80% confidence bands the entire null curve falls between the upper and lower limits even though the bands are narrower than the those with confidence 99%. In other words, a p-value greater than 0.20 indicates acceptance of the null hypotheses of degree 6 polynomial. The shape of the polynomial curve with $d = 6$ is consistent with the findings in Chaudhuri and Marron (1999) and Ruppert, Wand and Carroll (2003).

6 . Conclusions

We provide close forms of confidence bands constructed from polynomial spline regression. Asymptotic properties have been established for equally spaced, nonadaptive selection of knots. Extension to adaptive design is infeasible, as Härdle, Marron and Yang (1997) had shown that adaptive knots selection could lead to inconsistency in L_∞ norm.

It is possible, however, to extend the constant band in Theorem 1 to unequally spaced deterministic knots subject to mesh constraints as in (2003). The linear band in Theorem 2 does not allow such direct extension. This is one of the two reasons that the constant band remains viable despite the fact that the linear band has much better theoretical property and practical performance. The constant band is kept also for its simplicity. When implemented according to (4.5) with estimation on equally-spaced knots, the confidence limits at point x is the exact same as those at the nearest knot $t_{j(x)}$, so the constant band is in fact $(N+1)$ independently inflated confidence intervals. In contrast, the linear band has to be calibrated at each new point x. The confidence limits at x and the ones at $t_{j(x)}$ are different.

The main hurdle of generalizing our method to higher order splines is the inversion of the inner product matrix of B-spline basis, for which close form solutions exist in the case of linear spline with the aid of (4.10) and (4.11). Such close form formulae become unavailable for higher order splines.

Verbatim extension to multivariate regression is difficult for lack of sharp approximation as in (3.7). This limitation is also in Claeskens and Van Keilegom (2003) and Xia (1998). However, the univariate bands in this paper are still valuable for multivariate regression for the following reason. Semiparametric dimension reduction models such as the additive model, the partial linear model and the single index model provide various conduits of reducing multivariate nonparametric regression to some form of univariate smoothing. For instance, the components of additive model, the nonparametric component of partially linear model and the nonparametric link function of single index model, are all estimable via univariate smoothing. Therefore any new tool for univariate smoothing always brings fresh insights into high dimensional smoothing problems.

Appendix A: Proof of Theorem 1

A. 1 Preliminaries

Throughout Appendices A and B, we denote by the same letters c, C, any positive constants, without distinction in each case. The detailed proof is given at www.msu.edu/˜yangli/bandfull.pdf.

Lemma A.1 *Under Assumptions (A3) and (A4), there exists* $\alpha_0 > 0$ *such that the sequence* $\{D_n\} = \{n^{\alpha_0}\}$ *satisfies*

$$\frac{\log^2 n}{\sqrt{nh}} D_n \to 0, \sum_{n=1}^{\infty} D_n^{-(2+\delta)} < \infty, \frac{\sqrt{nh}}{D_n^{(1+\delta)}} \to 0, D_n^{-\delta} h^{-1/2} \to 0. \tag{A.1}$$

For such sequence $\{D_n\}, P\{\omega \,|\exists N(\omega), \ni |\varepsilon_i| \le D_n, 1 \le i \le n, n > N(\omega)\} = 1.$

Denote the theoretical norms of the basis $c_{j,n} = \|b_{j,1}\|_2^2$ and $d_{j,n} = \|b_{j,2}\|_2^2$ as

$$c_{j,n} = \int_a^b I_j(x) f(x)\, dx, d_{j,n} = \int_a^b K^2\left(\frac{x - t_{j+1}}{h}\right) f(x)\, dx.$$

Lemma A.2 *Under Assumptions (A2) and (A3), as* $n \to \infty$,

$$c_{j,n} = f(t_j) h (1 + r_{j,n,1}), \langle b_{j,1}, b_{j',1}\rangle \equiv 0, j \ne j' \tag{A.2}$$

$$d_{j,n} = \frac{2}{3} f(t_{j+1}) h \begin{cases} 1 + r_{j,n,2} & j = 0, ..., N-1, \\ 1/2 + r_{j,n,2} & j = -1, N, \end{cases} \tag{A.3}$$

$$\langle b_{j,2}, b_{j',2}\rangle = \frac{1}{6} f(t_{j+1}) h \begin{cases} 1 + \tilde{r}_{j,n,2} & |j' - j| = 1, \\ 0 & |j' - j| > 1, \end{cases} \tag{A.4}$$

where

$$\max\nolimits_{0 \le j \le N} |r_{j,n,1}| + \max\nolimits_{-1 \le j \le N} \{|r_{j,n,2}| + |\tilde{r}_{j,n,2}|\} \le C\omega(f, h). \tag{A.5}$$

In particular,

$$\frac{1}{3} f(t_{j+1}) h \{1 - C\omega(f, h)\} \le d_{j,n} \le \frac{2}{3} f(t_{j+1}) h \{1 + C\omega(f, h)\}. \tag{A.6}$$

Proof of Lemma 3.1. For brevity, we give only the proof of (3.1) for $A_{n,1}$. Take any $j = 0, 1, ..., N$

$$\left|\|B_{j,1}\|_{2,n}^2 - 1\right| = \left|\sum\nolimits_{i=1}^n \xi_i\right|, \xi_i = \left\{B_{j,1}^2(X_i) - 1\right\} n^{-1}$$

with $E\xi_i = 0$ and for any $k \ge 2$, Minkowski's inequality implies that

$$E|\xi_i|^k = n^{-k} E\left|B_{j,1}^2(X_i) - 1\right|^k \le 2^{k-1} n^{-k} E\left[B_{j,1}^{2k}(X_i) + 1\right] \le \left\{\frac{2}{nh}\right\}^k C_0 h,$$

while (A.2) entails that $E\xi_i^2 \ge n^{-2} E\left[\frac{1}{2} B_{j,1}^4(X_i) - 1\right] \ge \{2/(nh)\}^2 C_1 h$. One can then find a constant $c > 0$ such that for $k > 2$, $E|\xi_i|^k \le (cn^{-1}h^{-1})^{k-2} k! E|\xi_i|^2$. Applying Bernstein's inequality, one has $P\left\{|\sum_{i=1}^n \xi_i| \ge \delta \log^{1/2}(n)(nh)^{-1/2}\right\} \le 2n^{-3}$ for large enough $\delta > 0$. Thus,

$$\sum\nolimits_{n=1}^{\infty} P\left\{\sup\nolimits_{0 \le j \le N} \left|\|B_{j,1}\|_{2,n}^2 - 1\right| \ge \delta \log^{1/2}(n)(nh)^{-1/2}\right\} < \infty$$

for such $\delta > 0$, then (3.1) follows. $\square$

A. 2 Proof of Theorem 1

In this section, we investigate the behavior of $\tilde{\varepsilon}_1(x)$ defined in (3.4). Since $\left\langle \mathbf{B}_{j',1}(\mathbf{X}), \mathbf{B}_{j,1}(\mathbf{X})\right\rangle_n = 0$ unless $j = j'$, $\tilde{\varepsilon}_1(x)$ can be written as $\tilde{\varepsilon}_1(x) = \sum_{j=0}^N \varepsilon_j^* B_{j,1}(x) \|B_{j,1}\|_{2,n}^{-2}$, in which $\varepsilon_j^* = \left\langle \mathbf{E}, \mathbf{B}_{j,1}(\mathbf{X})\right\rangle_n = n^{-1}\sum_{i=1}^n B_{j,1}(X_i)\sigma(X_i)\varepsilon_i$.

Lemma A.3 *Let $\hat{\varepsilon}_1(x) = \sum_{j=0}^N \varepsilon_j^* B_{j,1}(x), x \in [a,b]$, for $A_{n,1}$ defined in (3.1)*

$$|\tilde{\varepsilon}_1(x) - \hat{\varepsilon}_1(x)| \le A_{n,1}(1 - A_{n,1})^{-1} |\hat{\varepsilon}_1(x)|, x \in [a,b].$$

Thus, $\sup_{x\in[a,b]} |\tilde{\varepsilon}_1(x)|$ and $\sup_{x\in[a,b]} |\hat{\varepsilon}_1(x)|$ have the same asymptotic behavior.

Lemma A.4 *The pointwise variance of $\hat{\varepsilon}_1(x)$ is the function $\sigma_{n,1}^2(x)$ defined in (2.5) which satisfies with* $\sup_{x\in[a,b]} |r_{n,1}(x)| \to 0$

$$E\{\hat{\varepsilon}_1(x)\}^2 \equiv \sigma_{n,1}^2(x) = \frac{\sigma^2(x)}{f(x)\,nh}\{1 + r_{n,1}(x)\}, x \in [a,b]. \tag{A.7}$$

Lemma A.5 *Let the sequence $\{D_n\}$ satisfy (A.1), then as $n \to \infty$*

$$\left\|\hat{\varepsilon}_{n,1}(x) - \hat{\varepsilon}_{n,1}^D(x)\right\|_\infty = O\left(D_n^{-(1+\delta)}\sqrt{nh}\right) = o(1), \;\; w.\ p.\ 1,$$

where, for $x \in [a,b]$

$$\hat{\varepsilon}_{n,1}(x) = \sigma_{n,1}(x)^{-1}\sum_{j=0}^N B_{j,1}(x)\varepsilon_j^* = \sigma_{n,1}(x)^{-1}\sum_{j=0}^N B_{j,1}(x)\left(\varepsilon_j^* - E\varepsilon_j^*\right),$$
$$\hat{\varepsilon}_{n,1}^D(x) = \sigma_{n,1}(x)^{-1}\sum_{j=0}^N B_{j,1}(x)\left(\varepsilon_j^* - E\varepsilon_j^*\right) I_{\{|\varepsilon|<D_n\}}. \tag{A.8}$$

Proof. Notice that $E\varepsilon_j^* = E\{n^{-1}\sum_{i=1}^n B_{j,1}(X_i)\sigma(X_i)\varepsilon_i\} = 0$, then

$$\hat{\varepsilon}_{n,1}(x) = \left\{\sigma_{n,1}(x)\sqrt{n}c_{j(x),n}\right\}^{-1}\int\int I_{j(x)}(v)\sigma(v)\varepsilon dZ_n(v,\varepsilon)$$

according to the definition of $Z_n(v,\varepsilon)$ in (3.9). The truncated part $\hat{\varepsilon}_{n,1}^D(x)$ is defined in (A.8). The tail part $\hat{\varepsilon}_{n,1}(x) - \hat{\varepsilon}_{n,1}^D(x)$ is bounded uniformly over $[a,b]$ by

$$\sup_{x\in[a,b]}\left|\left\{\sigma_{n,1}(x)\sqrt{n}c_{j(x),n}\right\}^{-1}\int\int I_{j(x)}(v)\sigma(v)\varepsilon I_{\{|\varepsilon|\ge D_n\}}dZ_n(v,\varepsilon)\right|$$
$$\le \sup_{x\in[a,b]}\left|\left\{\sigma_{n,1}(x)c_{j(x),n}\right\}^{-1}\frac{1}{n}\sum_{i=1}^n I_{j(x)}(X_i)\sigma(X_i)\varepsilon_i I_{\{|\varepsilon_i|\ge D_n\}}\right| \tag{A.9}$$
$$+\sup_{x\in[a,b]}\left|\left\{\sigma_{n,1}(x)c_{j(x),n}\right\}^{-1}\int\int I_{j(x)}(v)\sigma(v)\varepsilon I_{\{|\varepsilon|\ge D_n\}})dF(v,\varepsilon)\right|. \tag{A.10}$$

By Lemma A.1, the term (A.9) is 0 almost surely. The term (A.10) is bounded by

$$\sup_{x\in[a,b]}\left\{\sigma_{n,1}(x)\,c_{j(x),n}\right\}^{-1}\int I_{j(x)}(v)\,\sigma(v)\,f(v)\left[\int|\varepsilon|\,I_{\{|\varepsilon|\geq D_n\}}dF(\varepsilon\,|v)\right]dv$$
$$\leq\sup_{x\in[a,b]}\left\{\sigma_{n,1}(x)\,c_{j(x),n}\right\}^{-1}\int I_{j(x)}(v)\,\sigma(v)\,f(v)\,dv\frac{M_\delta}{D_n^{1+\delta}}\leq C\frac{\sqrt{nh}}{D_n^{1+\delta}}.$$

The lemma follows immediately by the third condition in (A.1). □

Lemma A.6 *Define for* $x\in[a,b]$

$$\hat{\varepsilon}_{n,1}^{(0)}(x)=\left\{\sigma_{n,1}(x)\sqrt{n}c_{j(x),n}\right\}^{-1}\int\int I_{j(x)}(v)\,\sigma(v)\,\varepsilon I_{\{|\varepsilon|<D_n\}}dB\left\{M(v,\varepsilon)\right\}$$

then as $n\to\infty$

$$\sup_{x\in[a,b]}\left|\hat{\varepsilon}_{n,1}^{(0)}(x)-\hat{\varepsilon}_{n,1}^{D}(x)\right|=O\left(h^{-1/2}n^{-1/2}D_n\log^2 n\right)=o(1),\ \ w.\ p.\ 1.$$

Proof. First, $\left|\hat{\varepsilon}_{n,1}^{(0)}(x)-\hat{\varepsilon}_{n,1}^{D}(x)\right|$ can be written as

$$\left|\left\{\sigma_{n,1}(x)\sqrt{n}c_{j(x),n}\right\}^{-1}\int\int I_{j(x)}(v)\,\sigma(v)\,\varepsilon I_{\{|\varepsilon|<D_n\}}d\left[Z_n(v,\varepsilon)-B\left\{M(v,\varepsilon)\right\}\right]\right|,$$

which becomes the following via integration by parts

$$\left|\left\{\sigma_{n,1}(x)\sqrt{n}c_{j(x),n}\right\}^{-1}\int\int\left[Z_n(v,\varepsilon)-B\left\{M(v,\varepsilon)\right\}\right]d\left\{I_{j(x)}(v)\,\sigma(v)\,\varepsilon I_{\{|\varepsilon|<D_n\}}\right\}\right|$$
$$\leq\left\{\sigma_{n,1}(x)\sqrt{n}c_{j(x),n}\right\}^{-1}\int\int\left|Z_n(v,\varepsilon)-B\left\{M(v,\varepsilon)\right\}\right|d\left\{\varepsilon I_{\{|\varepsilon|<D_n\}}\right\}d\left\{I_{j(x)}(v)\,\sigma(v)\right\}.$$

Next, by Lemma A.4, the bounded variation of the function $\sigma(x)$ in Assumption (A2), the strong approximation result (3.10) and the first condition in (A.1), $\sup_{x\in[a,b]}\left|\hat{\varepsilon}_{n,1}^{(0)}(x)-\hat{\varepsilon}_{n,1}^{D}(x)\right|$ is bounded as

$$O\left\{(nh)^{1/2}n^{-1/2}h^{-1}\left(n^{-1/2}\log^2 n\right)D_n\right\}=O\left(n^{-1/2}h^{-1/2}D_n\log^2 n\right)=o(1)$$

with probability 1, thus completing the proof of the lemma. □

The next lemma finds a process $\hat{\varepsilon}_{n,1}^{(1)}(x)$ defined in terms of the 2-dimensional Brownian motion to approximate $\hat{\varepsilon}_{n,1}^{(0)}(x)$.

Lemma A.7 *Define for* $x\in[a,b]$

$$\hat{\varepsilon}_{n,1}^{(1)}(x)=\left\{\sigma_{n,1}(x)\sqrt{n}c_{j(x),n}\right\}^{-1}\int\int I_{j(x)}(v)\,\sigma(v)\,\varepsilon I_{\{|\varepsilon|<D_n\}}dW\left\{M(v,\varepsilon)\right\}$$

then as $n\to\infty$, $\left\|\hat{\varepsilon}_{n,1}^{(1)}(x)-\hat{\varepsilon}_{n,1}^{(0)}(x)\right\|_\infty=O\left(h^{1/2}D_n^{-(1+\delta)}\right)=o(1),\ \ w.\ p.\ 1.$

Proof. Based on the Rosenblatt transformation $M(x,\varepsilon)$ defined in (3.8), and $\frac{\partial M(x,\varepsilon)}{\partial(x,\varepsilon)} = f(x,\varepsilon)$, then the term $\left\|\hat{\varepsilon}_{n,1}^{(1)}(x) - \hat{\varepsilon}_{n,1}^{(0)}(x)\right\|_{\infty}$ is bounded by

$$\begin{aligned}
&\sup_{x\in[a,b]}\left|\left\{\sigma_{n,1}(x)\sqrt{n}c_{j(x),n}\right\}^{-1}\int\int I_{j(x)}(v)\,\sigma(v)\,|\varepsilon|\,I_{\{|\varepsilon|<D_n\}}dM(v,\varepsilon)\,W(1,1)\right| \\
&\leq \sup_{x\in[a,b]}\left\{\sigma_{n,1}(x)\sqrt{n}c_{j(x),n}\right\}^{-1}\int I_{j(x)}(v)\,\sigma(v)\,f(v)\,dv \\
&\times\left\{\int|\varepsilon|\,I_{\{|\varepsilon|<D_n\}}f_{\varepsilon|v}(\varepsilon|v)d\varepsilon\right\}|W(1,1)| \\
&\leq C\left(\frac{\sqrt{nh}}{\sqrt{nh}}\right)h\frac{M_\delta}{D_n^{1+\delta}}|W(1,1)| = O\left(h^{1/2}D_n^{-(1+\delta)}\right) = o(1) \text{ w. p. } 1
\end{aligned}$$

The last step is obtained by applying the third condition in (A.1). □

The next lemma expresses the distribution of $\hat{\varepsilon}_{n,1}^{(1)}(x)$ in terms of 1-dimensional Brownian motion.

Lemma A.8 *The process $\hat{\varepsilon}_{n,1}^{(1)}(x)$ has the same distribution as*

$$\hat{\varepsilon}_{n,1}^{(2)}(x) = \left\{\sigma_{n,1}(x)\sqrt{n}c_{j(x),n}\right\}^{-1}\int I_{j(x)}(v)\,\sigma(v)\,s_n(v)\,f^{\frac{1}{2}}(v)\,dW(v)\,, x\in[a,b]$$

where

$$s_n^2(v) = \int\varepsilon^2 I_{\{|\varepsilon|<D_n\}}f_{\varepsilon|v}(\varepsilon|v)d\varepsilon. \tag{A.11}$$

Proof. As $\operatorname{var}\left\{\hat{\varepsilon}_{n,1}^{(1)}(x)\right\}$ and $\operatorname{var}\left\{\hat{\varepsilon}_{n,1}^{(2)}(x)\right\}$ are exactly the same for any $x\in[a,b]$. Hence, the two Gaussian processes $\hat{\varepsilon}_n^{(1)}(x)$ and $\hat{\varepsilon}_n^{(2)}(x)$ have the same distribution according to Itô's Isometry Theorem. □

Lemma A.9 *Define for any $x\in[a,b]$*

$$\hat{\varepsilon}_{n,1}^{(3)}(x) = \left\{\sigma_{n,1}(x)\sqrt{n}c_{j(x),n}\right\}^{-1}\int I_{j(x)}(v)\,\sigma(v)\,f^{\frac{1}{2}}(v)\,dW(v)$$

then as $n\to\infty$, $\left\|\hat{\varepsilon}_{n,1}^{(2)}(x) - \hat{\varepsilon}_{n,1}^{(3)}(x)\right\|_{\infty} = O\left(D_n^{-\delta}h^{-1/2}\right) = o(1)$ w. p. 1.

Proof. By the fourth condition in (A.1) , $\sup_{x\in[a,b]}\left|\hat{\varepsilon}_{n,1}^{(2)}(x) - \hat{\varepsilon}_{n,1}^{(3)}(x)\right|$ is almost surely bounded by

$$\begin{aligned}
&\sup_{v\in[a,b]}\left|s_n^2(v)-1\right|\sup_{x\in[a,b]}\left|\sigma_{n,1}^{-1}(x)\,c_{j(x),n}^{-1}n^{-1/2}\int I_{j(x)}(v)\,\sigma(v)\,f^{\frac{1}{2}}(v)\,dW(v)\right| \\
&= O\left(D_n^{-\delta}h^{-1/2}\right) = o(1)\,.
\end{aligned}$$
□

Lemma A.10 *The process* $\hat{\varepsilon}_{n,1}^{(3)}(x)$ *is a Gaussian process with mean* 0*, variance* 1*, and covariance* $\operatorname{cov}\left\{\hat{\varepsilon}_{n,1}^{(3)}(x), \hat{\varepsilon}_{n,1}^{(3)}(y)\right\} = \delta_{j(x),j(y)}, \forall x, y \in [a, b]$.

Proof. Itô's Isometry Theorem and (A.7) implies the lemma result. □

Proof of Proposition 3.1. The proof follows immediately from Lemmas A.3, A.5, A.6, A.7, A.8, A.9 and A.10. □

Proof of Theorem 1. It is clear from Proposition 3.1 that the Gaussian process $U(x)$ consists of $(N+1)$ i.i.d. standard normal variables $U(t_0), ..., U(t_N)$, hence Theorem 4 implies that as $n \to \infty$

$$P\left\{\sup_{x\in[a,b]} |U(x)| \leq \tau/a_{N+1} + b_{N+1}\right\} \to \exp\left(-2e^{-\tau}\right).$$

By letting $\tau = -\log\left\{-\frac{1}{2}\log(1-\alpha)\right\}$, and using the definition of a_{N+1} and b_{N+1}, we obtain

$$\lim_{n\to\infty} P\left[\sup_{x\in[a,b]} |U(x)| \leq -\log\left\{-\frac{1}{2}\log(1-\alpha)\right\}\{2\log(N+1)\}^{-1/2}\right.$$
$$\left.+\{2\log(N+1)\}^{1/2} - \frac{1}{2}\{2\log(N+1)\}^{-1/2}\{\log\log(N+1)+\log 4\pi\}\right] = 1-\alpha.$$

Replacing $U(x)$ with $\sigma_{n,1}(x)^{-1}\tilde{\varepsilon}_1(x)$ (Proposition 3.1), and the definition of d_n in (2.8) entail that

$$\lim_{n\to\infty} P\left[\sup_{x\in[a,b]} \left|\sigma_{n,1}(x)^{-1}\tilde{\varepsilon}_1(x)\right| \leq \{2\log(N+1)\}^{1/2} d_n\right] = 1-\alpha.$$

As (3.5) entails that $\sqrt{\log(N+1)/(nh)}\,\|\tilde{m}_1(x) - m(x)\|_\infty = o_p(1)$. Thus according to (3.3)

$$\begin{aligned}
&\lim_{n\to\infty} P\left[m(x) \in \hat{m}_1(x) \pm \sigma_{n,1}(x)\{2\log(N+1)\}^{1/2} d_n, \forall x \in [a,b]\right] \\
=\ &\lim_{n\to\infty} P\left[\{2\log(N+1)\}^{-1/2} d_n^{-1} \sup_{x\in[a,b]} \sigma_{n,1}^{-1}(x)\left|\tilde{\varepsilon}_1(x) + \tilde{m}_1(x) - m(x)\right| \leq 1\right] \\
=\ &\lim_{n\to\infty} P\left[\{2\log(N+1)\}^{-1/2} d_n^{-1} \sup_{x\in[a,b]} \sigma_{n,1}^{-1}(x)\left|\tilde{\varepsilon}_1(x)\right| \leq 1\right] = 1-\alpha. \quad \square
\end{aligned}$$

Appendix B: Proof of Theorem 2

B. 1 Preliminaries

In this subsection we examine matrices used in (2.9) of Theorem 2. In what follows, we use $|T|$ to denote the maximal absolute value of any matrix T, and M_{N+2} is the tridiagonal matrix as defined in (4.9).

Lemma B.1 *The inner product matrix* V *of the B-spline basis* $\{B_{j,2}(x)\}_{j=-1}^{N}$ *defined as* $V = (v_{j'j})_{j,j'=-1}^{N} = (\langle B_{j',2}, B_{j,2}\rangle)_{j,j'=-1}^{N}$ *,which has the following decomposition*

$$V = M_{N+2} + (\tilde{v}_{j'j})_{j,j'=-1}^{N} = M_{N+2} + \tilde{V}$$

where $\tilde{v}_{j'j} \equiv 0$ *if* $|j-j'| \geq 1$, *and* $\left|\tilde{V}\right| \leq C\omega(f,h)$.

Proof. By (A.3), (A.4) and (A.5), the inner product of $\langle b_{j',2}, b_{j,2}\rangle$ can be replaced by $\frac{1}{6}f(t_{j+1})h$ if $|j'-j| = 1$, and $\frac{1}{3}f(t_{j+1})h$ or $\frac{2}{3}f(t_{j+1})h$ when $j' = j$, plus some uniformly infinitesimal differences dominated by $\omega(f,h)$. Then based on the definition of $B_{j,2}(x)$, the lemma follows immediately. □

The next lemma shows that multiplication by M_{N+2} behaves similarly to multiplication by a constant.

Lemma B.2 *Given matrix* $\Omega = M_{N+2} + \Gamma$, *in which* $\Gamma = \left(\gamma_{jj'}\right)_{j,j'=-1}^{N}$ *satisfies* $\gamma_{jj'} \equiv 0$ *if* $|j-j'| \geq 1$ *and* $|\Gamma| \xrightarrow{p} 0$. *Then there exist constants* $c, C > 0$ *independent of* n *and* Γ, *such that in probability*

$$c\,|\boldsymbol{\xi}| \leq |\Omega\boldsymbol{\xi}| \leq C\,|\boldsymbol{\xi}|\,, C^{-1}\,|\boldsymbol{\xi}| \leq \left|\Omega^{-1}\boldsymbol{\xi}\right| \leq c^{-1}\,|\boldsymbol{\xi}|\,, \forall \boldsymbol{\xi} \in R^{N+2}. \tag{B.1}$$

Proof. Since each row of M_{N+2} has diagonal element equal to 1, and one or two nonzero off-diagonal terms whose total absolute values do not exceed $1/\sqrt{2}$, hence $\left(1 - 1/\sqrt{2} - 3\,|\Gamma|\right)|\boldsymbol{\xi}| \leq |\Omega\boldsymbol{\xi}| \leq 3\,(1 + |\Gamma|)\,|\boldsymbol{\xi}|$,which entails the first inequality of (B.1), and the second one follows by switching the roles of $\boldsymbol{\xi}$ and $\Omega\boldsymbol{\xi}$. □

As an application of Lemma B.2, consider the matrix $S = V^{-1}$ defined in (2.4). Let $\tilde{\boldsymbol{\xi}}_{j'} = \{\mathrm{sgn}\,(s_{j'j})\}_{j=-1}^{N}$, then there exists a positive C_s such that

$$\sum\nolimits_{j=-1}^{N} |s_{j'j}| \leq \left|S\tilde{\boldsymbol{\xi}}_{j'}\right| \leq C_s\left|\tilde{\boldsymbol{\xi}}_{j'}\right| = C_s, \forall j' = -1, 0, ..., N. \tag{B.2}$$

The matrix S in the construction of the confidence band can not be computed exactly as it involves the unknown density $f(x)$. We approximate S by the inverse of M_{N+2}, with a simpler, distribution-free form in (4.9). This approximation is uniform for S_j in (2.4) and Ξ_j in (4.8) as well.

Lemma B.3 *As* $n \to \infty$, $\left|M_{N+2}^{-1} - S\right| \to 0$ *and* $\max_{0\leq j\leq N}|\Xi_j - S_j| \to 0$.

Proof. By definition, $M_{N+2}M_{N+2}^{-1} = I = VS = \left(M_{N+2} + \tilde{V}\right)S$. Denote by e_i the unit vector with i-th element 1, then applying Lemma B.2 with $\Omega = M_{N+2}$,

$$\begin{aligned} & c\left|M_{N+2}^{-1} - S\right| = c\max\nolimits_{i=1}^{N+2}\left|\left(M_{N+2}^{-1} - S\right)e_i\right| \\ \leq\; & \max\nolimits_{i=1}^{N+2}\left|M_{N+2}\left(M_{N+2}^{-1} - S\right)e_i\right| \leq \left|\tilde{V}\right|\left(\left|M_{N+2}^{-1} - S\right| + \left|M_{N+2}^{-1}\right|\right) \end{aligned}$$

Since Lemma B.1 makes $\left|\tilde{V}\right| \leq C\omega(f,h)$, as $n \to \infty$, $\left|M_{N+2}^{-1} - S\right| = O\{\omega(f,h)\} \to 0$. Now by definition of submatrices S_j and Ξ_j, $\max_{0\leq j\leq N}|\Xi_j - S_j| \leq \left|M_{N+2}^{-1} - S\right|$, the lemma follows. $\square$

B. 2 Variance calculation

We now examine the asymptotic behavior of

$$\tilde{\varepsilon}_2(x) = \mathrm{Proj}_{G_n^{(0)}} \mathbf{E} = \sum\nolimits_{j=-1}^{N} \tilde{a}_j B_{j,2}(x), x \in [a,b] \tag{B.3}$$

where the coefficient vector $\tilde{\mathbf{a}} = (\tilde{a}_{-1}, ..., \tilde{a}_N)^T$ are solutions to the normal equations

$$\left(\langle B_{j,2}, B_{j',2}\rangle_n\right)_{j,j'=-1}^{N} (\tilde{a}_j)_{j=-1}^{N} = \left(n^{-1}\sum\nolimits_{i=1}^{n} B_{j,2}(X_i)\sigma(X_i)\varepsilon_i\right)_{j=-1}^{N}.$$

In other words

$$\tilde{\mathbf{a}} = (\tilde{a}_j)_{j=-1}^{N} = \left(V + \tilde{B}\right)^{-1} \left(n^{-1}\sum\nolimits_{i=1}^{n} B_{j,2}(X_i)\sigma(X_i)\varepsilon_i\right)_{j=-1}^{N}, \tag{B.4}$$

where $\left|\tilde{B}\right| \leq A_{n,2} = O_p\left(n^{-1/2}h^{-1/2}\log^{1/2}(n)\right)$ by (3.2).

Now define $\hat{a}_j$'s by replacing $\left(V + \tilde{B}\right)^{-1}$ with $V^{-1} = S$ in above formula, i.e.

$$\hat{\mathbf{a}} = (\hat{a}_j)_{j=-1}^{N} = \left(\sum\nolimits_{j=-1}^{N} s_{j'j} n^{-1}\sum\nolimits_{i=1}^{n} B_{j,2}(X_i)\sigma(X_i)\varepsilon_i\right)_{j=-1}^{N} \tag{B.5}$$

and define for $x \in [a,b]$

$$\hat{\varepsilon}_2(x) = \sum_{j=-1}^{N} \hat{a}_j B_{j,2}(x) = \sum_{j,j'=-1}^{N} s_{j'j} n^{-1}\sum\nolimits_{i=1}^{n} B_{j,2}(X_i)\sigma(X_i)\varepsilon_i B_{j',2}(x). \tag{B.6}$$

The next lemma is a special case of the unconditional version of (6.2) in (2003).

Lemma B.4 *The pointwise variance of $\hat{\varepsilon}_2(x)$ is the function $\sigma_{n,2}^2(x)$ defined in (2.5), which satisfies*

$$E\left\{\hat{\varepsilon}_2^2(x)\right\} \equiv \sigma_{n,2}^2(x) = \frac{3\sigma^2(x)}{2f(x)nh}\mathbf{\Delta}^T(x) S_{j(x)} \mathbf{\Delta}(x)\{1 + r_{n,2}(x)\}$$

with $\sup_{x\in[a,b]}|r_{n,2}(x)| \to 0$, $j(x)$ in (2.2), $\mathbf{\Delta}(x)$ in (4.7) and matrix S_j in (2.4). Consequently, there exist $0 < c_\sigma < C_\sigma$ such that for large enough n

$$c_\sigma (nh)^{-1/2} \leq \sigma_{n,2}(x) \leq C_\sigma (nh)^{-1/2}, \forall x \in [a,b]. \tag{B.7}$$

Proof. See Wang and Yang (2006). $\square$

B. 3 Proof of Theorem 2

The next several lemmas are for the proof of Proposition 3.2.

Lemma B.5 *Define for $x \in [a,b]$*

$$\begin{aligned}\hat{\varepsilon}_{n,2}(x) &= \sigma_{n,2}^{-1}(x)\,\hat{\varepsilon}_2(x) = \sigma_{n,2}^{-1}(x)\sum\nolimits_{j=-1}^{N} \hat{a}_{j'}B_{j',2}(x),\\ \hat{\varepsilon}_{n,2}^{D}(x) &= \sigma_{n,2}^{-1}(x)\sum\nolimits_{j=-1}^{N} \hat{a}_{j'}B_{j',2}(x)\, I_{\{|\varepsilon|<D_n\}}.\end{aligned}$$

where D_n satisfies (A.1). Then with probability 1

$$\left\|\hat{\varepsilon}_{n,2}(x) - \hat{\varepsilon}_{n,2}^{D}(x)\right\|_{\infty} = O\left(n^{1/2}h^{1/2}D_n^{-(1+\delta)}\right) = o(1).$$

Proof. Since obviously $E\hat{\varepsilon}_{n,2}(x) = 0$, $\forall x \in [a,b]$,

$$\hat{\varepsilon}_{n,2}(x) = \sigma_{n,2}^{-1}(x)\, n^{-1/2} \sum_{j'=j(x)-1}^{j(x)} B_{j',2}(x) \sum_{j=-1}^{N} s_{j'j} \int\int B_{j,2}(v)\,\sigma(v)\,\varepsilon\, dZ_n(v,\varepsilon)$$

where $Z_n(x,\varepsilon)$ is defined in (3.9). The technical proof is very similar to Lemma A.5, except that we employ (B.2) to deal with $\sum_{j=-1}^{N} s_{j'j}$. The same order is also achieved. □

Lemma B.6 *Let M be the Rosenblatt transformation given in (3.8) and define*

$$\hat{\varepsilon}_{n,2}^{(0)}(x) = \frac{1}{\sqrt{n}\sigma_{n,2}(x)} \sum_{j'j=-1}^{N} B_{j',2}(x)\, s_{j'j} \int\int B_{j,2}(v)\,\sigma(v)\,\varepsilon I_{\{|\varepsilon|<D_n\}} dB\left\{M(v,\varepsilon)\right\}$$

for $x \in [a,b]$. Then as $n \to \infty$

$$\sup_{x\in[a,b]} \left|\hat{\varepsilon}_{n,2}^{(0)}(x) - \hat{\varepsilon}_{n,2}^{D}(x)\right| = O\left(n^{-1/2}h^{-1/2}D_n \log^2 n\right) = o(1),\ \ w.\ p.\ 1.$$

Proof. See Lemma A.6. □

Lemma B.7 *Define for $x \in [a,b]$*

$$\hat{\varepsilon}_{n,2}^{(1)}(x) = \frac{\sigma_{n,2}^{-1}(x)}{\sqrt{n}} \sum_{j'j=-1}^{N} B_{j',2}(x)\, s_{j'j} \int\int B_{j,2}(v)\,\sigma(v)\,\varepsilon I_{\{|\varepsilon|<D_n\}} dW\left\{M(v,\varepsilon)\right\},$$

then as $n \to \infty$

$$\sup_{x\in[a,b]} \left|\hat{\varepsilon}_{n,2}^{(1)}(x) - \hat{\varepsilon}_{n,2}^{(0)}(x)\right| = O\left(h^{1/2}D_n^{-(1+\delta)}\right) = o(1),\ \ w.\ p.\ 1.$$

Lemma B.8 *The process* $\hat{\varepsilon}^{(1)}_{n,2}(x), x \in [a,b]$ *has the same distribution as*

$$\hat{\varepsilon}^{(2)}_{n,2}(x) = \sigma^{-1}_{n,2}(x)\, n^{-1/2} \sum\nolimits_{j'j=-1}^{N} B_{j',2}(x)\, s_{j'j} \iint B_{j,2}(v)\,\sigma(v)\, s_n(v)\, f^{\frac{1}{2}}(v)\, dW(v)$$

for $x \in [a,b]$, *where* $s_n^2(v)$ *is as defined in (A.11).*

Proof. Similar to that of Lemma A.8, see Wang and Yang (2006) for details. □

Lemma B.9 *Define for any* $x \in [a,b]$

$$\hat{\varepsilon}^{(3)}_{n,2}(x) = \frac{1}{\sqrt{n}\sigma_{n,2}(x)} \sum_{j'j=-1}^{N} B_{j',2}(x)\, s_{j'j} \int B_{j,2}(v)\,\sigma(v)\, f^{\frac{1}{2}}(v)\, dW(v)$$

then $\operatorname{var}\left\{\hat{\varepsilon}^{(3)}_{n,2}(x)\right\} \equiv 1, \forall x \in [a,b]$, *and as* $n \to \infty$

$$\left\|\hat{\varepsilon}^{(2)}_{n,2}(x) - \hat{\varepsilon}^{(3)}_{n,2}(x)\right\|_{\infty} = O\left(h^{-1/2} D_n^{-\delta}\right) = o(1)\,,\ \ w.\ p.\ 1.$$

Proof. Using (A.1) in the last step, the term $\sup_{x\in[a,b]}\left|\hat{\varepsilon}^{(2)}_{n,2}(x) - \hat{\varepsilon}^{(3)}_{n,2}(x)\right|$ is bounded by

$$\sup_{x\in[a,b]}\left|1 - s_n^2(x)\right| \sup_{x\in[a,b]}\left\{\frac{\sigma^{-1}_{n,2}(x)}{\sqrt{n}} \sum_{j'j=-1}^{N} B_{j',2}(x)\left|s_{j'j}\right| \int B_{j,2}(v)\,\sigma(v)\, f^{\frac{1}{2}}(v)\, dW(v)\right\}$$

$$\leq M_\delta D_n^{-\delta} h^{1/2} C \left|\int \sigma(v)\, f^{\frac{1}{2}}(v)\, dW(v)\right| = O\left(h^{-1/2} D_n^{-\delta}\right) = o(1) \text{ w. p. 1.}$$

Meanwhile, directly from (2.6) and (2.5), for any $x \in [a,b]$

$$\operatorname{var}\left\{\hat{\varepsilon}^{(3)}_{n,2}(x)\right\} = E\left\{\frac{\sigma^{-1}_{n,2}(x)}{\sqrt{n}} \sum_{j'j=-1}^{N} B_{j',2}(x)\, s_{j'j} \int B_{j,2}(v)\,\sigma(v)\, f^{\frac{1}{2}}(v)\, dW(v)\right\}^2 = 1.$$

□

Now define for any $j' = -1, ..., N$ and $x \in [a,b]$, the functions

$$\zeta_{j'}(x) = n^{-1/2}\sigma^{-1}_{n,2}(x)\, B_{j',2}(x)\,, \tilde{\boldsymbol{\zeta}}(x) = \left(\zeta_{j(x)-1}(x)\,, \zeta_{j(x)}(x)\right)^T$$

and the random vector $\boldsymbol{\Lambda} = (\Lambda_{-1}, \Lambda_0, \ldots, \Lambda_N)^T$ where

$$\Lambda_{j'} = \sum_{j=-1}^{N} s_{j'j} \int\int B_{j,2}(v)\,\sigma(v)\, f^{\frac{1}{2}}(v)\, dW(v)\,.$$

Then $\boldsymbol{\Lambda} \sim \mathbf{N}\left(\mathbf{0}, S\sum S\right)$ as $E\Lambda_{j'} = 0, \forall j' = -1, .., N$, and the covariance is $E\Lambda_{j'}\Lambda_{l'} = \sum_{j,l=-1}^{N} s_{j'j}\sigma_{jl}s_{ll'}$, for any $j', l' = -1, .., N$, and σ_{jl} is defined in (2.6). Notice that

$$\hat{\varepsilon}^{(3)}_{n,2}(x) \equiv \sum_{j'=j(x)-1,j(x)} \zeta_{j'}(x)\,\Lambda_{j'} = \tilde{\boldsymbol{\zeta}}(x)^T \boldsymbol{\Lambda}_{j(x)}, \boldsymbol{\Lambda}_j = \left(\Lambda_{j-1}, \Lambda_j\right)^T, j = 0, ..., N.$$

Since Lemma B.9 states that the term $\hat{\varepsilon}_{n,2}^{(3)}(x)$ always has variance 1, it means

$$\hat{\varepsilon}_{n,2}^{(3)}(x) = \frac{\tilde{\boldsymbol{\zeta}}(x)^T \boldsymbol{\Lambda}_{j(x)}}{\sqrt{\tilde{\boldsymbol{\zeta}}(x)^T \left\{\mathrm{cov}\left(\boldsymbol{\Lambda}_{j(x)}\right)\right\} \tilde{\boldsymbol{\zeta}}(x)}}. \tag{B.8}$$

Lemma B.10 *For any given* $0 < \alpha < 1$, *one has*

$$\liminf_{n\to\infty} P\left(\sup_{x\in[a,b]} \left|\hat{\varepsilon}_{n,2}(x)\right| \leq \left[2\left\{\log(N+1) - \log\alpha\right\}\right]^{1/2}\right) \geq 1-\alpha. \tag{B.9}$$

PROOF. Define for any $0 \leq j \leq N, Q_j = \boldsymbol{\Lambda}_j^T \left\{\mathrm{cov}\left(\boldsymbol{\Lambda}_j\right)\right\}^{-1} \boldsymbol{\Lambda}_j$. Result 4.7 (a), page 140 of Johnson and Wichern (1992) ensures that Q_j is distributed as χ_2^2, hence

$$P\left[Q_j > 2\left\{\log(N+1) - \log\alpha\right\}\right] = \frac{\alpha}{N+1}, \forall 0 \leq j \leq N.$$

Then (B.8) and the Maximization Lemma of Johnson and Wichern (1992), page 66 ensures that

$$\left\{\hat{\varepsilon}_{n,2}^{(3)}(x)\right\}^2 = \frac{\left|\tilde{\boldsymbol{\zeta}}(x)^T \boldsymbol{\Lambda}_{j(x)}\right|^2}{\tilde{\boldsymbol{\zeta}}(x)^T \left\{\mathrm{cov}\left(\boldsymbol{\Lambda}_{j(x)}\right)\right\} \tilde{\boldsymbol{\zeta}}(x)} \leq \boldsymbol{\Lambda}_{j(x)}^T \left\{\mathrm{cov}\left(\boldsymbol{\Lambda}_{j(x)}\right)\right\}^{-1} \boldsymbol{\Lambda}_{j(x)} = Q_{j(x)},$$

for any $x \in [a,b]$. One has therefore $\sup_{x\in[a,b]} \left|\hat{\varepsilon}_{n,2}^{(3)}(x)\right|^2 \leq \max_{0\leq j\leq N}\left\{Q_j\right\}$ and

$$\begin{aligned} & P\left[\sup_{x\in[a,b]} \left|\hat{\varepsilon}_{n,2}^{(3)}(x)\right|^2 \leq 2\left\{\log(N+1) - \log\alpha\right\}\right] \\ \geq\ & P\left[\max_{0\leq j\leq N}\left\{Q_j\right\} > 2\left\{\log(N+1) - \log\alpha\right\}\right] \geq 1-\alpha. \end{aligned}$$

Now (B.9) follows from Lemmas B.5, B.6 B.7, B.8, B.9. □

Lemma B.11

$$\left|\sup_{x\in[a,b]}\left|\frac{\hat{\varepsilon}_2(x)}{\sigma_{n,2}(x)}\right| - \sup_{x\in[a,b]}\left|\frac{\tilde{\varepsilon}_2(x)}{\sigma_{n,2}(x)}\right|\right| = O_p\left(\sqrt{\frac{\log n}{nh}}\right) = o_p(1).$$

PROOF. Recall the definition for $\tilde{\mathbf{a}} = \left(\tilde{a}_{-1}, \tilde{a}_0, ..., \tilde{a}_N\right)^T$ and $\hat{\mathbf{a}} = \left(\hat{a}_{-1}, \hat{a}_0, \ldots, \hat{a}_N\right)^T$ in (B.4) and (B.5), one has $\left(V + \tilde{B}\right)\tilde{\mathbf{a}} = V\hat{\mathbf{a}}$. Based on Lemma B.2 and (3.2), there exists a constant c such that $c\left|\hat{\mathbf{a}} - \tilde{\mathbf{a}}\right| \leq \left|V\left(\hat{\mathbf{a}} - \tilde{\mathbf{a}}\right)\right| = \left|\tilde{B}\tilde{\mathbf{a}}\right| \leq A_{n,2}\left(\left|\hat{\mathbf{a}} - \tilde{\mathbf{a}}\right| + \left|\hat{\mathbf{a}}\right|\right)$, it implies that $\left|\hat{\mathbf{a}} - \tilde{\mathbf{a}}\right| \leq \frac{A_{n,2}}{c-A_{n,2}}\left|\hat{\mathbf{a}}\right|$. From the definitions of $\tilde{\varepsilon}_2(x)$ in (B.3) and $\hat{\varepsilon}_2(x)$ in (B.6), plus (B.7) and (A.6), as $n \to \infty$

$$\sup_{x\in[a,b]}\left|\frac{\hat{\varepsilon}_2(x)}{\sigma_{n,2}(x)} - \frac{\tilde{\varepsilon}_2(x)}{\sigma_{n,2}(x)}\right| \leq \sup_{x\in[a,b]}\left|\sum\nolimits_{j=-1}^{N} \frac{\left|\hat{\mathbf{a}} - \tilde{\mathbf{a}}\right| B_{j,2}(x)}{\sigma_{n,2}(x)}\right| \leq Cn^{1/2}\frac{A_{n,2}}{c-A_{n,2}}\left|\hat{\mathbf{a}}\right|.$$

Use (A.6) again, it implies that as $n \to \infty$

$$\sup_{x\in[a,b]} \left| \frac{\hat{\varepsilon}_2(x)}{\sigma_{n,2}(x)} \right| \geq \frac{\sqrt{nh}}{C_\sigma} \sup_{x\in[a,b]} \left| \hat{\mathbf{a}}\mathbf{B}_2^T(x) \right| \geq C\sqrt{n}\,|\hat{\mathbf{a}}|$$

where $\mathbf{B}_2(x) = \{B_{-1,2}(x), ..., B_{N,2}(x)\}^T$, $\mathbf{b}_2(x) = \{b_{-1,2}(x), ..., b_{N,2}(x)\}^T$.

Then the desired result follows, i.e.

$$\sup_{x\in[a,b]} \left| \frac{\hat{\varepsilon}_2(x)}{\sigma_{n,2}(x)} - \frac{\tilde{\varepsilon}_2(x)}{\sigma_{n,2}(x)} \right| \leq C\frac{A_{n,2}}{c - A_{n,2}} \sup_{x\in[a,b]} \left| \frac{\hat{\varepsilon}_2(x)}{\sigma_{n,2}(x)} \right| = O_p\left(\sqrt{\frac{\log n}{nh}}\right). \qquad \square$$

Proof of Proposition 3.2. It follows from Lemma B.10 and Lemma B.11 automatically. $\square$

Proof of Theorem 2. Now (3.5) entails that

$$(nh)^{-1/2}\sqrt{\log(N+1)}\,\|\tilde{m}_2(x) - m(x)\|_\infty = O_p\left\{(nh)^{-1/2}\sqrt{\log(n)}h^2\right\} = o_p(1).$$

Applying (3.6) in Proposition 3.2

$$\begin{aligned}
&\liminf_{n\to\infty} P\left[m(x) \in \hat{m}_2(x) \pm \sigma_{n,2}(x)\{2\log(N+1) - 2\log\alpha\}^{1/2}, \forall x \in [a,b]\right] \\
= &\liminf_{n\to\infty} P\left[\sup_{x\in[a,b]} \sigma_{n,2}^{-1}(x)\,|\tilde{\varepsilon}_2(x) + \tilde{m}_2(x) - m(x)| \leq \{2\log(N+1) - 2\log\alpha\}^{1/2}\right] \\
= &\liminf_{n\to\infty} P\left[\sup_{x\in[a,b]} \left|\frac{\tilde{\varepsilon}_2(x)}{\sigma_{n,2}(x)}\right| \leq \{2\log(N+1) - 2\log\alpha\}^{1/2}\right] \geq 1 - \alpha. \qquad \square
\end{aligned}$$

Acknowledgments

This research is part of the first author's dissertation work at Michigan State University under the supervision of the second author, and has been supported in part by NSF grants DMS 0405330, BCS 0308420 and SES 0127722.

References

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095.

Chaudhuri, P. and Marron, J.S. (1999). SiZer for exploration of structures in curves. *J. Amer. Statist. Assoc.* **94** 807-823.

Claeskens, G. and Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31** 1852–1884.

de Boor, C. (2001). *A Practical Guide to Splines.* Springer-Verlag, New York.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall, London.

Gantmacher, F. R. and Krein, M. G. (1960). *Oszillationsmatrizen, Oszillationskerne und kleine Schwingungen mechanischer Systeme.* Akademie-Verlag, Berlin.

Hall, P. and Titterington, D. M. (1988). On confidence bands in nonparametric density estimation and regression. *J. Multi. Analys.* **27** 228–254.

Härdle, W. (1989). Asymptotic maximal deviation of M-smoothers. *J. Multi. Analys.* **29** 163–179.

Härdle, W. , Hlávka, Z. and Klinke, S. (2000). *XploRe Application Guide.* Springer-Verlag, Berlin.

Härdle, W., Marron, J. S. and Yang, L. (1997). Discussion of "Polynomial splines and their tensor products in extended linear modeling" by Stone et. al. *Ann. Statist.* **25** 1443-1450.

Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *Ann. Statist.* **31** 1600-1635.

Johnson, R. A. and Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis.* Prentice-Hall, New Jersey.

Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes.* Springer-Verlag, New York.

Mack, Y. P. and Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z.Wahrscheinlichkeitstheorie verm. Gebiete* **61** 405-415.

Rosenblatt, M. (1976). On the maximal deviation of k-dimensional density estimates. *Ann. Prob.* **4** 1009-1015.

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003). *Semiparametric Regression.* Cambridge University Press, Cambridge.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

Stone, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *Ann. Statist.* **22** 118–184.

Tusnády, G. (1977). A remark on the approximation of the sample df in the multidimensional case. *Periodica Mathematica Hungarica* **8** 53-55.

Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *J. Roy. Statist. Assoc. B* **60** 797–811.

Wang, J. and Yang, L. (2006). Polynomial Spline Confidence Bands for Regression Curves. http://www.msu.edu/˜ yangli/bandfull.pdf.

Zhang, F. (1999). *Matrix Theory. Basic Results and Techniques.* Springer-Verlag, New York.

Zhou, S., Shen, X. and Wolfe, D. A. (1998). Local asymptotics of regression splines and confidence regions. *Ann. Statist.* **26** 1760-1782.

Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, IL 60607, USA

E-mail: wangjing@math.uic.edu

Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA

E-mail: yang@stt.msu.edu

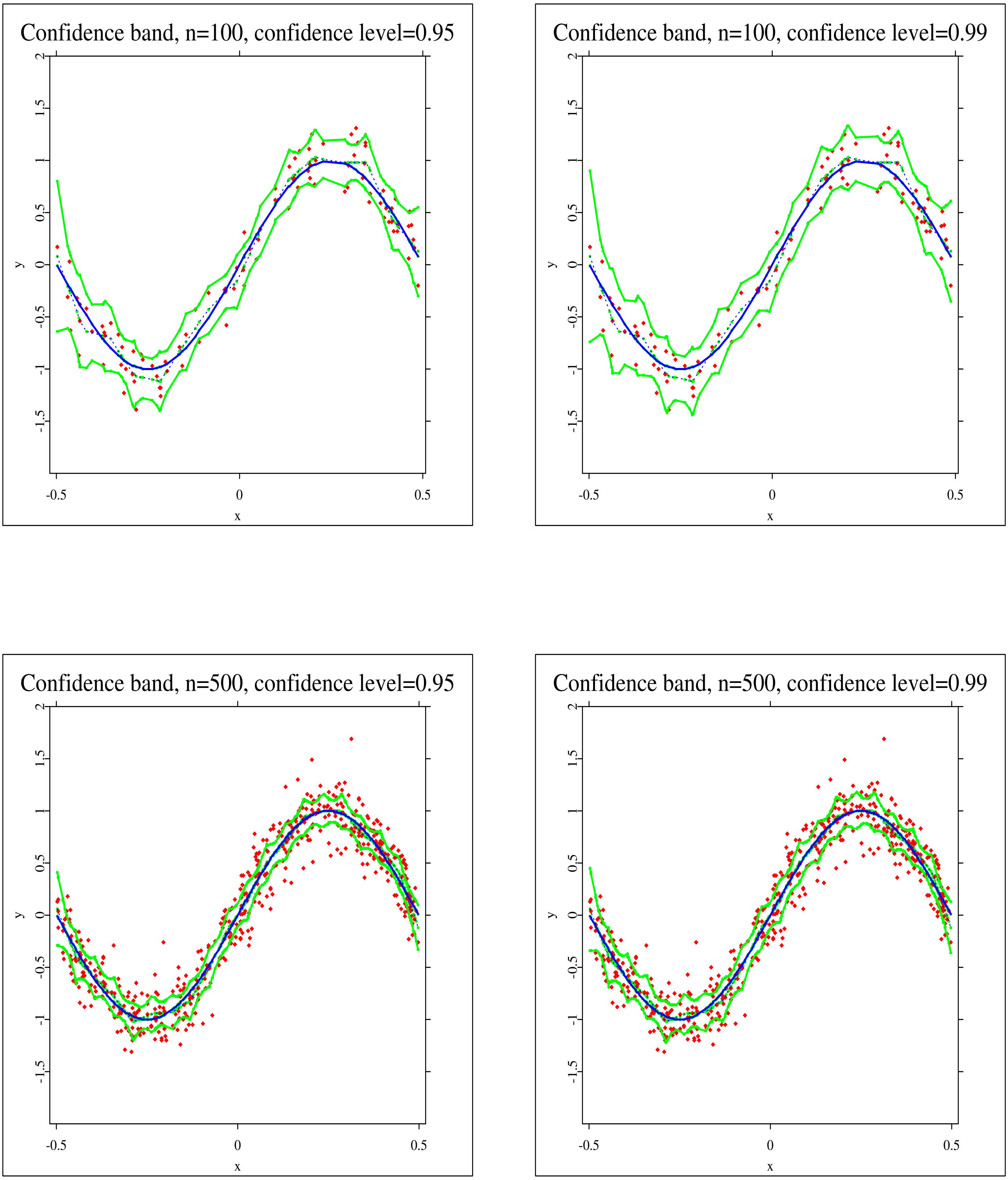

Figure 1: Plots of confidence bands (thick solid curves), the linear spline estimator $\hat{m}_2(x)$ (dashed curve), the true function $m(x) = \sin(2\pi x)$ (thin solid curve), and the data scatter plots. The bands are computed from (4.6) with opt $= 1$.

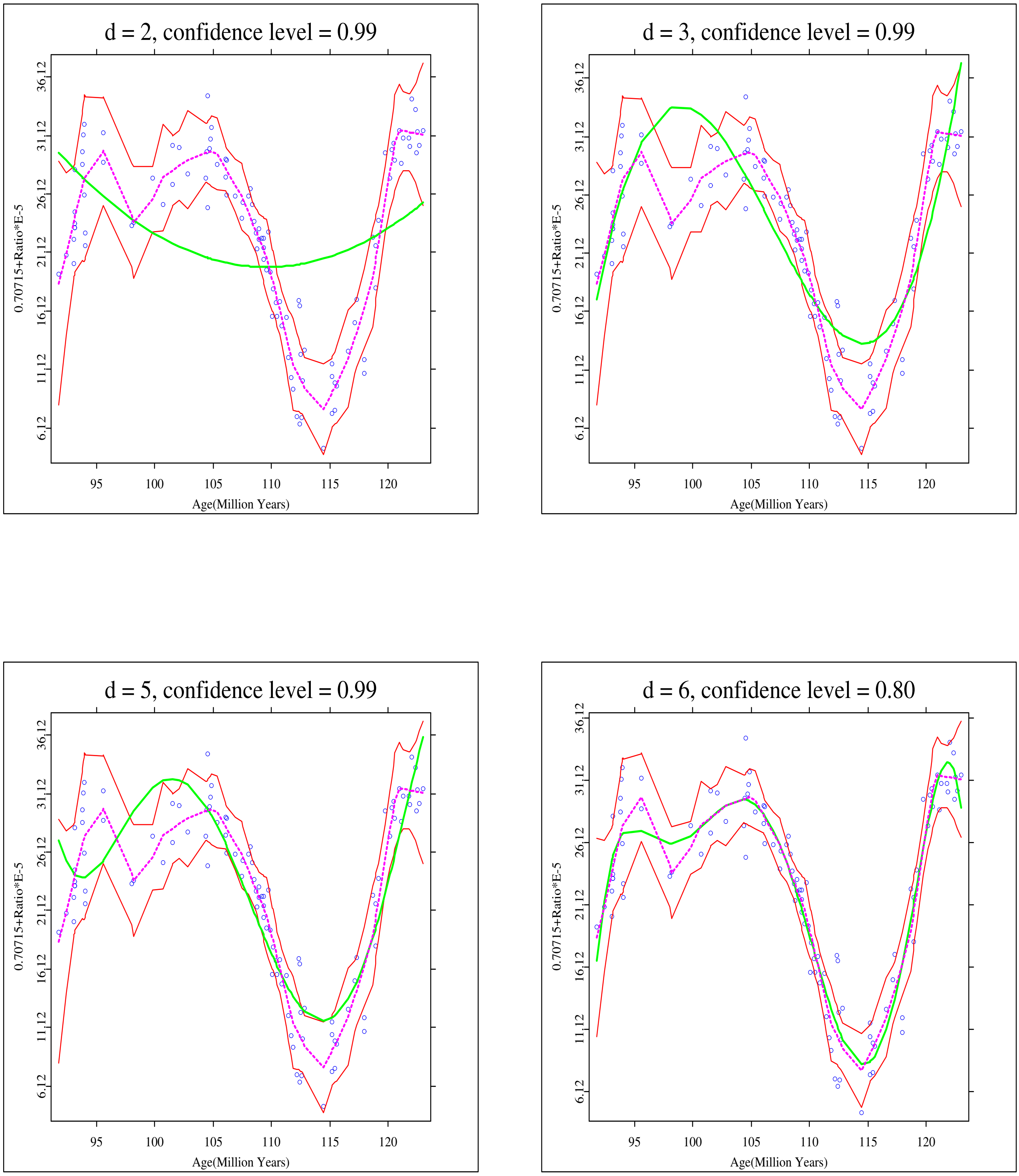

Figure 2: Plots of null hypothesis curves of $H_0 : m(x) = \sum_{k=1}^{d} a_k x^k, d = 2, 3, 5, 6$ (solid line), linear confidence bands (upper and lower thin lines), the linear spline estimator (dotted line) and the data (circle).